Universal Storage Consistency of DASD and Virtual Tape

Jim Erdahl
U.S.Bank

August, 14, 2013
Session Number 13848

AGENDA

- Context – mainframe tape and DLm
- Motivation for DLm8000
- DLm8000 implementation
- GDDR automation for DLm8000

Connotation of “tape”

- Tape no longer refers to physical tape media
- A different Tier of Storage (non-SMS)
- Characteristics
  - Serial access
  - Sequential stream of data
  - Special catalog and retentions controls (CA-1)
  - Lower cost than DASD
  - Limited number of drives (256 per VTE)
  - “No” space allocation constructs

~~(5GB x 255 volumes = 1,275GB of compressed data)~~

Characterization of mainframe tape data

- Archive / Retrieval
- Operational files
- Backup / Recovery

Profile at US Bank

- At least 60% of overall tape data is archive, including HSM ML2, CA View reports, and CMOD statements / images; this data is retrieved extensively
- No more than 25% of tape data is associated with backups, including image copies and archive logs

Tape data is critical!

DLm Origin / Primer

Origin	Key Component	Role
Bus-Tech MDL	VTEs	Tape drive emulation, FICON interface, compression, tape library portal
EMC NAS (Celerra / CLARiiON)	Data Movers	File System sharing over an IP network / replication
	SATA disks	Data storage

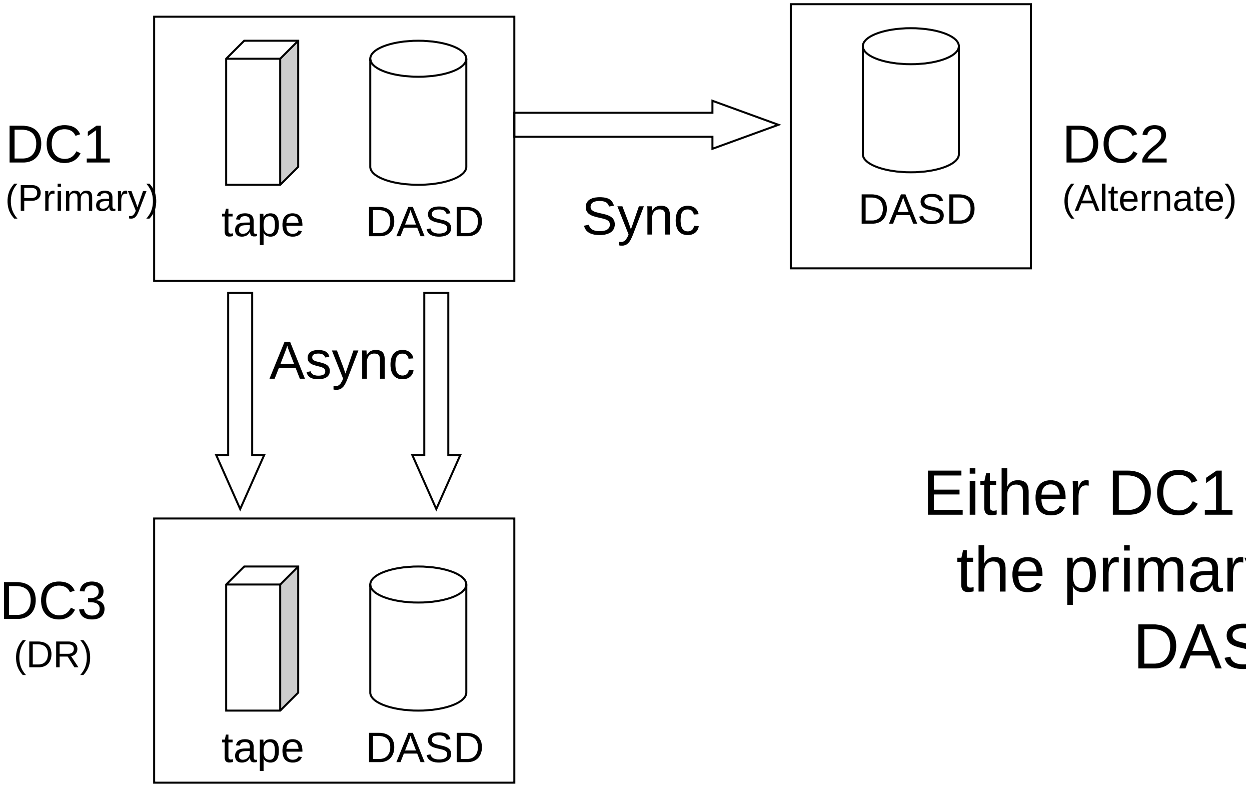
- Tape files reside in File Systems; the name of the file is the Volser
- Multiple file systems are defined within a tape library to provide for sufficient volsers and capacity headroom
- A single tape library is typically created per Sysplex and is mounted on specified VTE drives

DLm Requirements

DLm implemented in 2008 for a technology refresh in conjunction with a Data Center migration, providing:

- Consistently good performance
- Low maintenance
- High scalability and redundancy

Original US Bank Storage Configuration / Replication



Either DC1 or DC2 is the primary site for DASD

DLm Experience at US Bank

Notable metrics

- Over 50,000 mounts per day
- 99.9% of all tape mounts fulfilled in less than 1 second
- 4.5 to 1 compression
- Over 720 terabytes of usable capacity
- Peak host read / write rate of 1,200 megabytes / second
- Async replication to DC3 (DR Site) within several minutes
- Tape is becoming more critical

If life is so good, what is the motivation for DLm8000?

Problem #1 - Disaster Recovery Scenario The "missing tape" problem at DC3 (out-of-region DR site)

- Tape replication lags behind DASD replication (RPO measured in minutes versus seconds)
- In an out-of-region disaster declaration, tens and maybe hundreds of tape files closed immediately before the "Disaster" have not completely replicated
- But these files are defined in catalogs replicated on DASD (TMS, ICF catalogs, HSM CDSs, IMS RECON, DB2 BSDS, etc.)
- Hence, there are critical data inconsistencies between tape and DASD at DC3 (DR Site)

Problem #1 (continued)

During a Disaster Recovery at DC3 ….

1. What if HSM recalls fail because of missing ML2 tape data?
2. What if the DBAs cannot perform database recoveries because archive logs are missing?
3. What if business and customer data archived to tape is missing?
4. How does this impact overall recovery time (RTO)?
5. Are Disaster Recovery capabilities adequate if tapes are missing?

Problem #2 – Local Resiliency Scenario Local DASD resiliency is not sufficient

- Three site DASD configuration with synchronous replication between DC1 and DC2 and asynchronous replication to DC3. Two site tape configuration with asynchronous replication from DC1 to DC3.

- In the event of a DASD catastrophe at the primary site, a local DASD failover is performed non-disruptively with zero data loss, and asynchronous replication is re-instated to DC3.

But, what if a catastrophe occurs to tape storage at DC1?

- Mainframe processing will come to a grinding halt.
- A disruptive recovery can be performed at DC3, but this is a last resort.

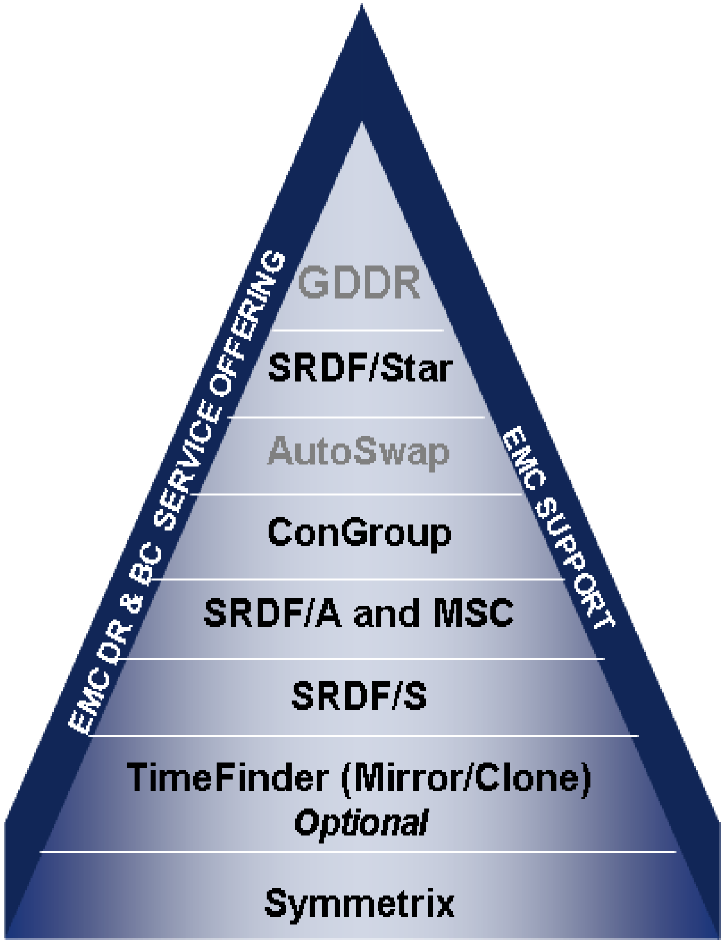
REMBER: Tape is becoming more essential.

What are the options?

- To solve problems #1 and #2, why not convert all tape allocations to DASD?
- The cost is prohibitive, but
- Space allocation is the impediment
  - SMS and Data Classes are not adequate
  - Massive JCL conversion is required to stipulate space allocation

- To solve problems #1 and #2, why not synchronize tape and DASD replication / failover?

EMC has the technologies, but …



EMC Foundation Technologies

1. Can DLm support a Symmetrix backend?

Yes

2. Can SRDF/S and SRDF/A handle enormous tape workloads without impacts?

Yes

3. Can we achieve “universal data consistency” between DASD and tape at DC2 and DC3?

Yes

4. Can GDDR manage a SRDF/STAR configuration with Autoswap which includes the tape infrastructure?

Yes

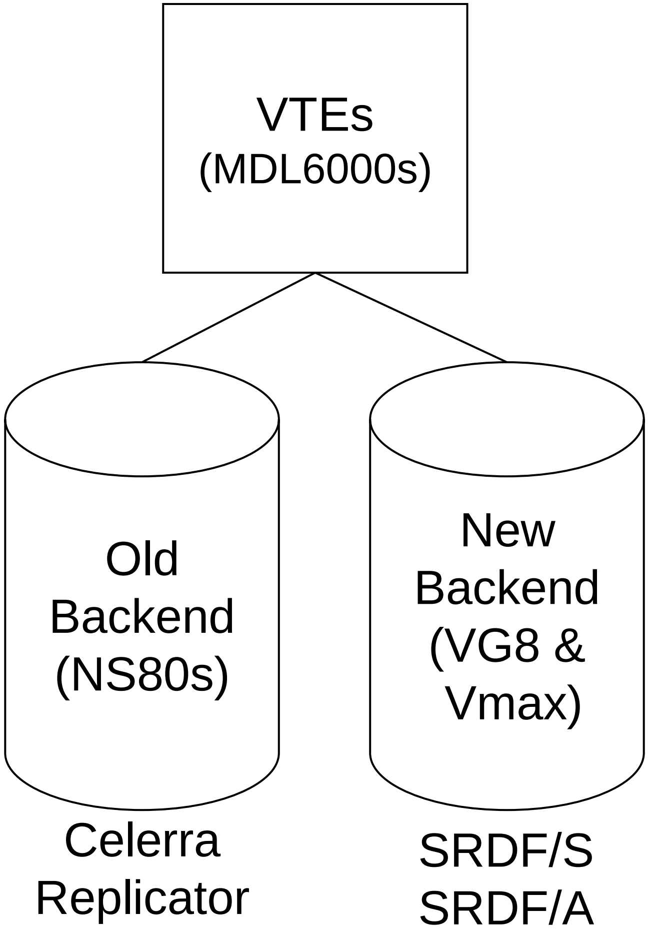
Creation of DLm8000

- US Bank tape resiliency requirements articulated in executive briefings and engineering round tables
- EMC solicited requirements from other customers and created a business case
- EMC performed a proof of concept validating overall functionality including failover processes
- EMC built a full scale configuration, based on US Bank's capacity requirements; performed final validation of replication, performance, and failover capabilities
- GDDR automation designed and developed with significant collaboration across product divisions
- US Bank began implementation in June, 2012

DLm8000 – what is under the hood?

Product Line	Key Component	Role	Management Interface
MDL 6000	VTEs	Tape drive emulation, FICON interface, compression, tape library portal	ACPs
VNX VG8	Data Movers	File System sharing over an IP network	Control Stations
Vmax	SATA FBA drives / SRDF	Data storage / replication	Gatekeepers

Migration to the DLm8000



1. Configured new backend at all three sites
2. Setup SRDF/A and SRDF/S
3. Defined File Systems on new backend
4. Partitioned “old” and “new” file systems with DLm storage classes
5. Updated Scratch synonyms to control scratch allocations by storage class
6. Deployed outboard DLm migration utility to copy tape files
7. Maintained dual replication from old backend and new backend during the migration
8. Incorporated tape into GDDR along with DASD (ConGroup, MSC, STAR, etc.).

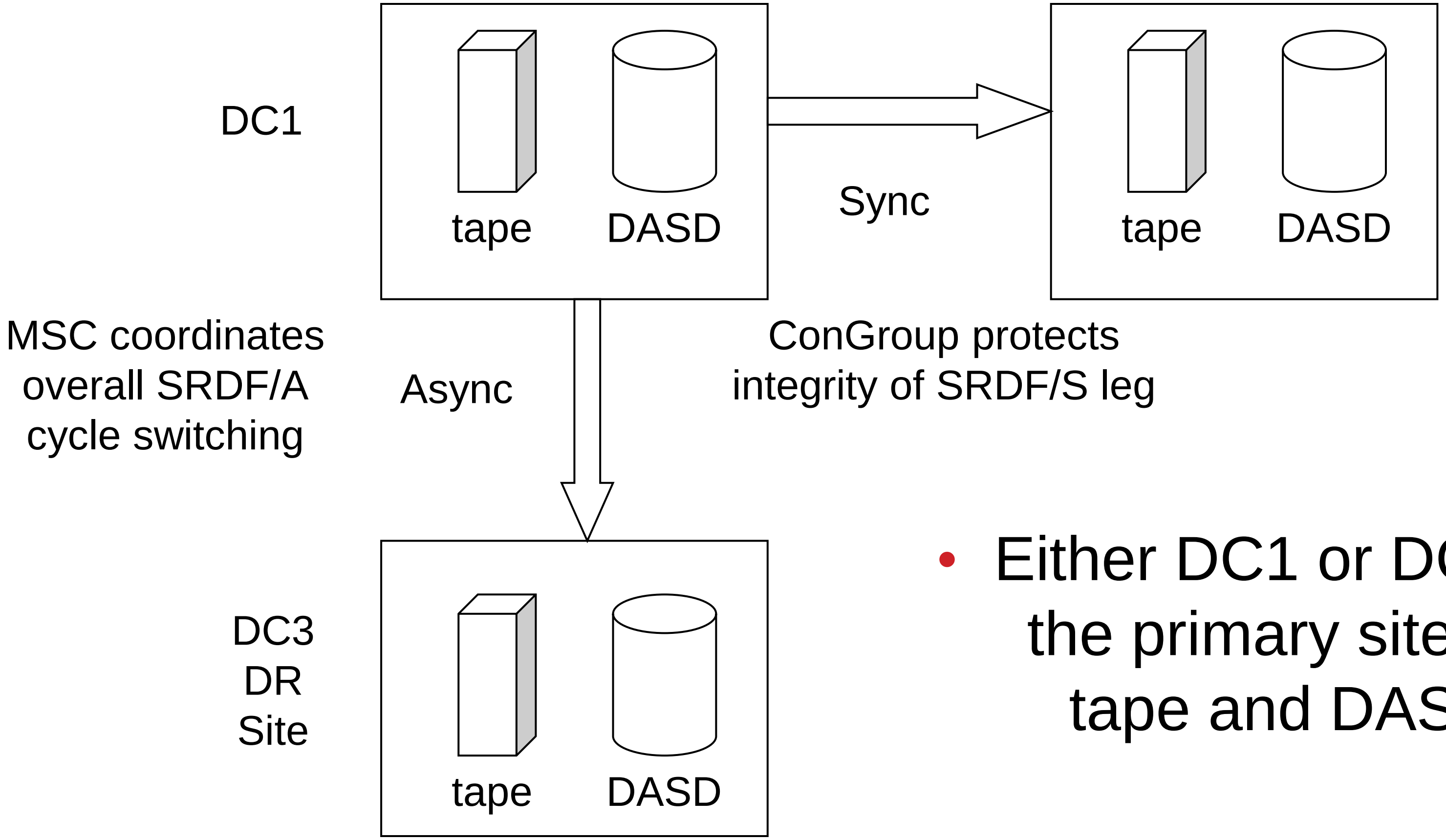
DLm Migration Utility – Process Overview

- Storage administrator creates a list of Volsers and starts the migration utility script on a VTE, with Volser list and parameter specifications

- For each Volser, the migration utility performs the following actions:
  - creates the target file in the specified storage class
  - locks the source and target files
  - copies the source to the target and performs a checksum
  - renames the source and target files so that the target file is “live”
  - unlocks the files

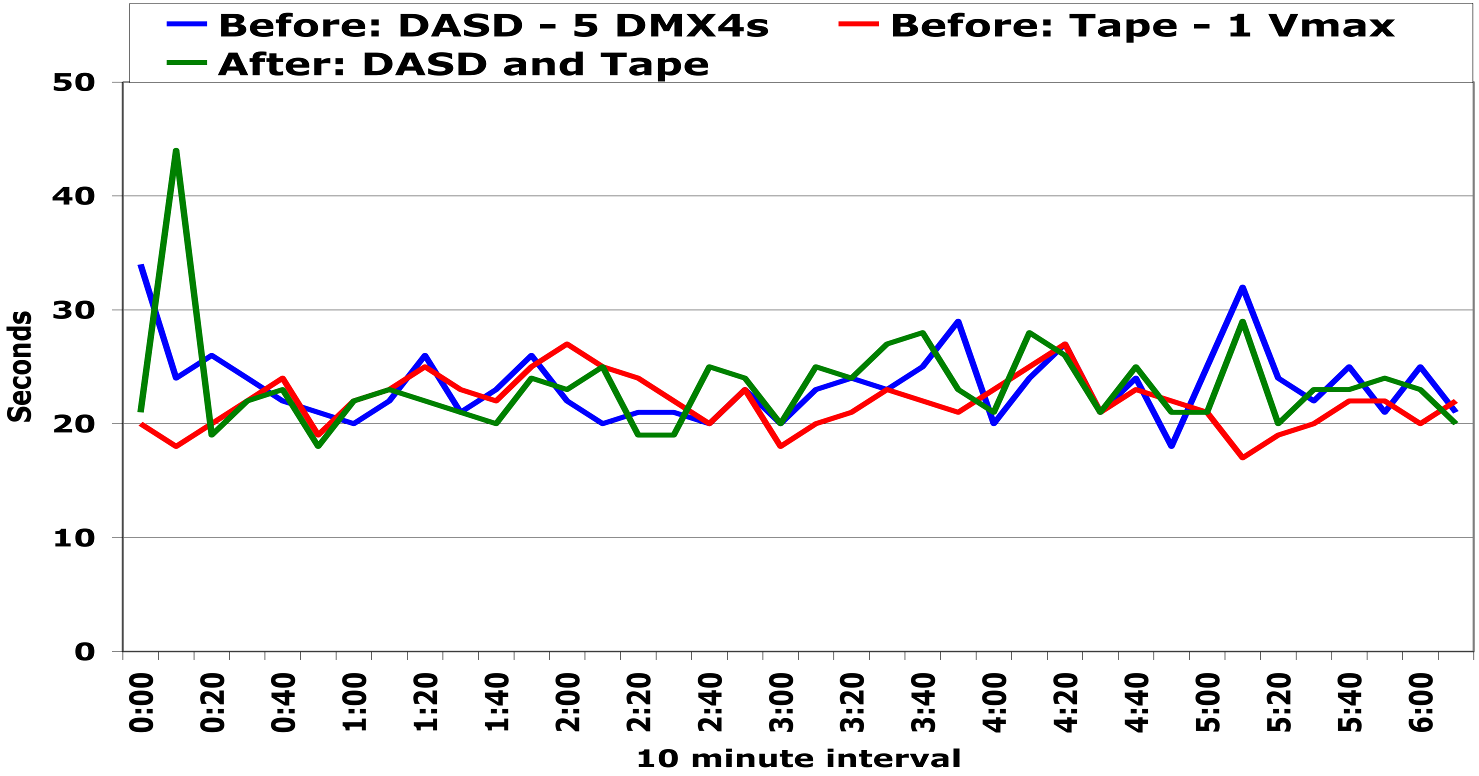
- An average throughput of 25 megabytes per second was achieved per VTE script without impacts to SRDF replication and no host impacts other than an occasional mount delay while a Volser is locked (max copy time < 80 minutes for 20GB tape)

Final US Bank Storage Configuration / Replication


DC1
tape
DASD
Sync
tape
DASD
DC2
MSC coordinates overall SRDF/A cycle switching
Async
ConGroup protects integrity of SRDF/S leg
DC3 DR Site
tape
DASD

- Either DC1 or DC2 is the primary site for tape and DASD

SRDF/A RPO - Before and After MSC Consolidation



How do we monitor / control all of this?

Geographically Dispersed Disaster Restart (GDDR) scripts were updated for DLm8000.

- **Recover At DC3**
- **Test from BCV's at DC3 and DC2**
- **Restart SRDF/A replication to DC3**
- **Restart SRDF/S replication between DC1 and DC2**
- **Planned Autoswap between DC1 and DC2**
- **Un-Planned Autoswap between DC1 and DC2**

GDDR automation for a planned storage swap between DC1 and DC2 (high level steps)

Once tape workload is quiesced, GDDR script is initiated …

1. Tape drives varied offline
2. Tape configuration disabled at “swap from” site
3. DASD Autoswap and SRDF swap (R1s not ready, R2s ready and R/W)
4. Failover of Data Movers to “swap to” site
5. VTEs started at “swap to” site
6. Tape drives varied online and tape processing resumes
7. Recovery / resumption of SRDF/S, SRDF/A, and STAR

Unplanned storage swap between DC1 and DC2

- Loss of access to CKD devices triggers DASD Autoswap and SRDF swap
- In-flight tape processing fails - no tape “Autoswap” functionality
- Failed in-flight tape jobs need to be re-started after the tape swap
- No data loss for closed tape files (or sync points)
- Note: a tape infrastructure issue does not trigger an unplanned swap
- GDDR recovery script is automatically triggered after an unplanned swap

GDDR script to recover from unplanned swap (high level steps)

1. Tape drives are varied offline
2. Failover of Data Movers to “swap to” site
3. VTEs started at “swap to” site
4. Tape drives varied online and tape processing resumes
5. Recovery / resumption of SRDF/A
6. Recovery / resumption of SRDF/S and STAR initiated manually

GDDR Value at US Bank

- Provides sophisticated automation for monitoring and management
- Handles coordination of comprehensive EMC software and hardware stack
- Provides recovery for critical events such as unplanned swaps and SRDF/A outages
- Minimizes requirements for internally developed automation and procedures
- Indispensable for a multi-site, high availability configuration

DLm8000 Value Proposition – High Availability

- Robust storage platform
- Synchronous and asynchronous replication without impacts
- Universal data consistency between tape and DASD at DC2 and DC3, enabled with ConGroup and MSC
- GDDR automation to manage overall storage replication, failover, and recovery

Universal Storage Consistency of DASD and Virtual Tape

Jim Erdahl
U.S.Bank

August, 14, 2013
Session Number 13848

Complete your sessions evaluation online at SHARE.org/BostonEval

SHARE in Boston